

**LTD : Lecture Textométrique Différentielle  
de corpus de versions parallèles**  
*MkAlign, iTrameur, Trameur*



**Humanités numériques  
et texte littéraire traduit**

Éditer et analyser des corpus de versions parallèles

24 septembre 2019

MEDIAT, salle des conférences - Domaine universitaire St-Martin-d'Hères



Journée d'étude organisée par Filippo Fonio, Emanuela Nanni et Pascale Roux, le 24 septembre 2019, Université Grenoble Alpes

**Maria ZIMINA-POIROT, CLILLAC-ARP**



# Définition de textométrie

*La textométrie regroupe l'ensemble des méthodes quantitatives permettant d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur l'ensemble des unités textuelles d'un corpus*

***Statistique textuelle, L. Lebart et A. Salem (1994)***

Dans un **contexte multilingue**, cette approche offre des perspectives de recherches prometteuses pour de multiples dimensions d'analyse de corpus dans des langues différentes (alignement automatique, extraction de ressources traductionnelles, navigation intertextuelle, synthèse de l'information, etc.).

# Lecture Textométrique Différentielle (LTD)

« Nous appelons *Lecture Textométrique Différentielle (LTD)* une stratégie de lecture de textes comparables appuyée par l’affichage synchrone des résultats de leur analyse textométrique parallèle (différences ou similarités caractéristiques). Les deux ensembles textuels sont affichés simultanément à l’écran pour faciliter les comparaisons. »

« Lecture Textométrique Différentielle (LTD) de textes législatifs comparables de l’Union européenne », S. Patin, M. Zimina, S. Fleury, *JADT 2016*.

# Corpus 17 | 2017 :

## Segments phraséologiques et séquences textuelles : méthodologie et caractérisation

<https://journals.openedition.org/corpus/2868>



133 %



### Figure 9. Lecture Textométrique Différentielle (LTD) de « réaliser » dans D-EU-preamble et T-FR-disposition

#### Directive européenne (Préambule)

(4) Dans un souci de clarté et de plus grande lisibilité, il convient d'intégrer dans la présente directive la disposition sur l'application de la directive 79/1072/CEE, qui figurait jusqu'à présent dans la directive 2006/112/CE du Conseil du 28 novembre 2006 relative au système commun de taxe sur la valeur ajoutée [4].

(5) Étant donné que les objectifs de la présente directive ne peuvent pas être réalisés de manière suffisante par les États membres et peuvent donc, en raison des dimensions de l'action, être mieux réalisés au niveau communautaire, la Communauté peut prendre des mesures, conformément au principe de subsidiarité consacré à l'article 5 du traité. Conformément au principe de proportionnalité tel qu'énoncé audit article, la présente directive n'excède pas ce qui est nécessaire pour atteindre ces objectifs.

Agrandir Original (png, 16k)

#### Législation française (Dispositif)

intérieure de consommation sur les produits énergétiques [Dispositions déclarées non conformes à la Constitution par la décision du Conseil constitutionnel n° 2009-599 DC du 29 décembre 2009] ;

« - les quotes-parts de résultat sur opérations faites en commun ;

« - les charges financières afférentes aux immeubles d'exploitation ;

« - les dotations aux amortissements d'exploitation ;

« - les dotations aux provisions autres que les provisions techniques.

« VII. - Lorsque les plus-values de cession d'immobilisations ou de titres de nature à être comprises dans le chiffre d'affaires et la valeur ajoutée en application des I et III à VI sont réalisées l'année de création de l'entreprise, elles sont comprises dans le chiffre d'affaires et la valeur ajoutée retenus au titre de l'année suivante.

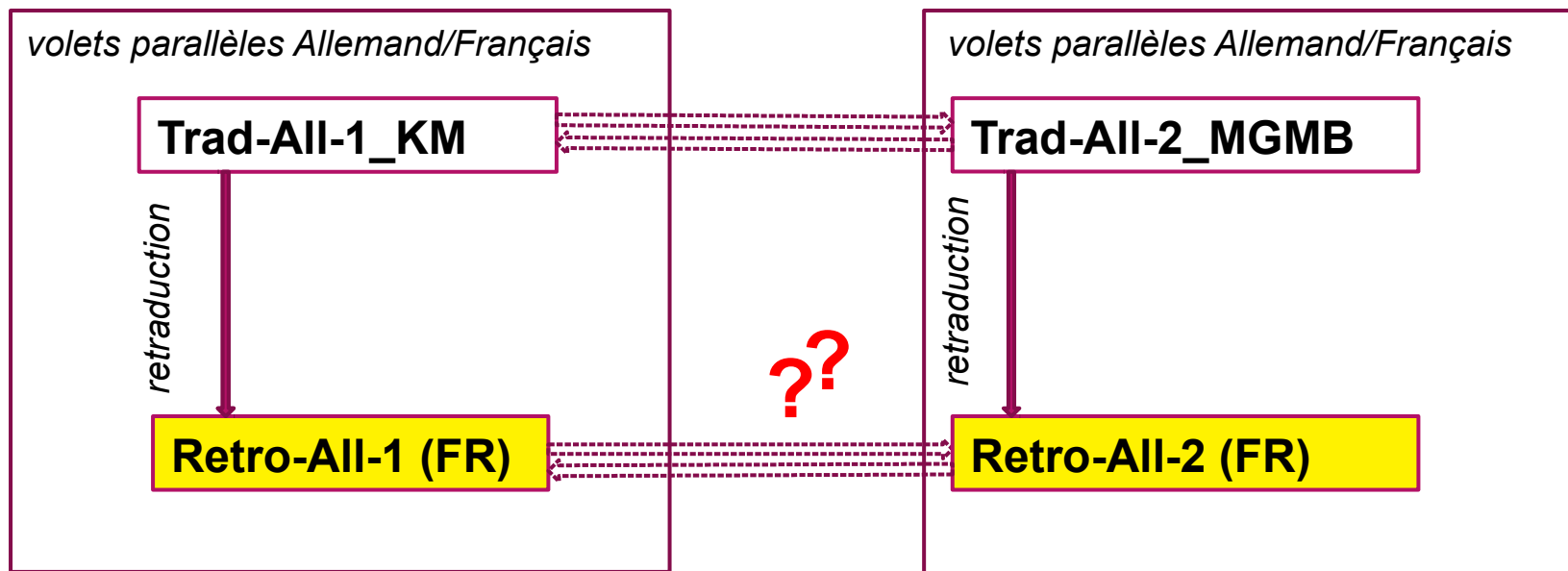
« Art. 1586 septies.-Le montant de la cotisation sur la valeur ajoutée des

C. Gledhill, S. Patin,  
M. Zimina

« Lexico-grammaire  
et textométrie :  
identification et  
visualisation de  
schémas lexico-  
grammaticaux  
caractéristiques dans  
deux corpus  
juridiques  
comparables en  
français »

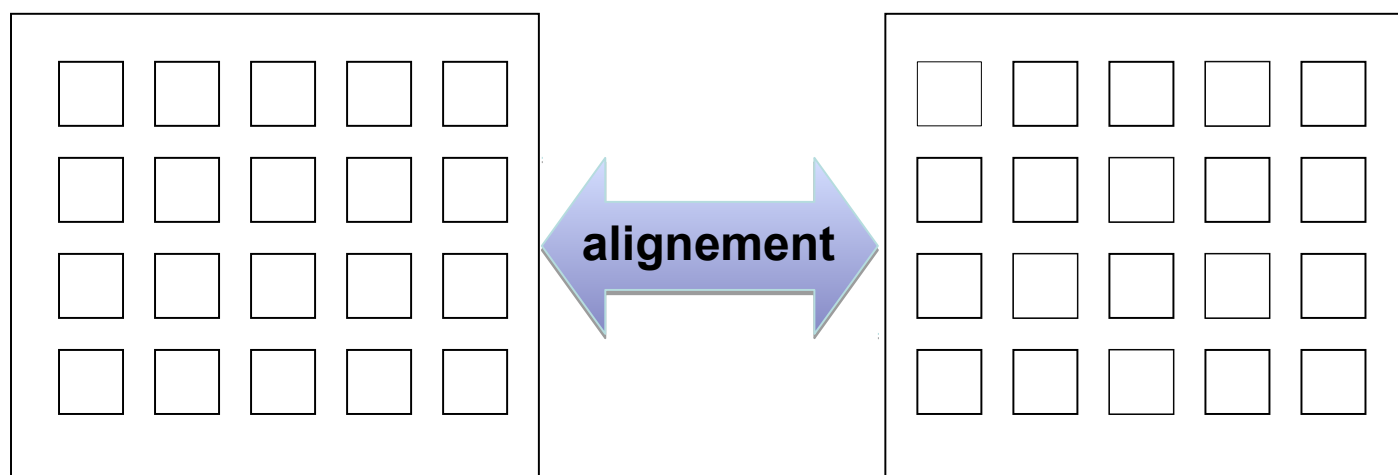
# Extrait du corpus MACE : Prose PROCHE AFRIQUE (ALL\_FR)

*produits simultanément*



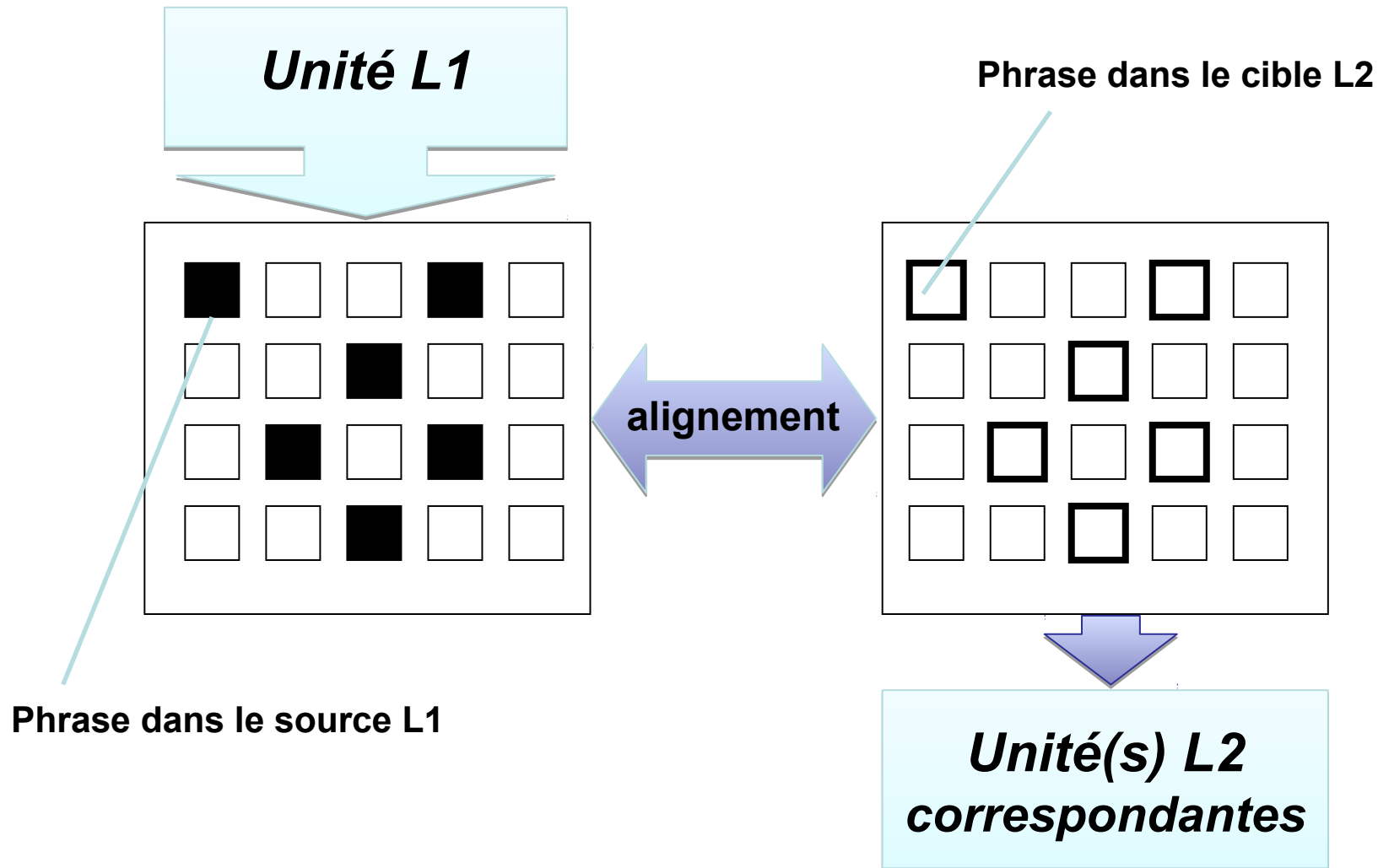
# 1. Bi-textes (corpus alignés) : démo *MkAlign*

## Corpus parallèles alignés



# Résonances textuelles

(Salem, Zimina, JADT 2004 ; S. Fleury : *MkAlign*)



# La méthode des spécificités

(Lebart et Salem, 1994. *Statistique textuelle*)

*PARTIES*

<i>Unités textuelles</i>			
		$K_{ij}$	$F_i$
		$t_j$	

Si l'effectif  $K_{ij}$  ne se situe pas dans les limites de ce que le **calcul probabiliste** permettait d'espérer, on calcule un *indice de spécificité* : *sur-emploi* ou *sous-emploi* de l'unité (spec.  $\pm XX$ )

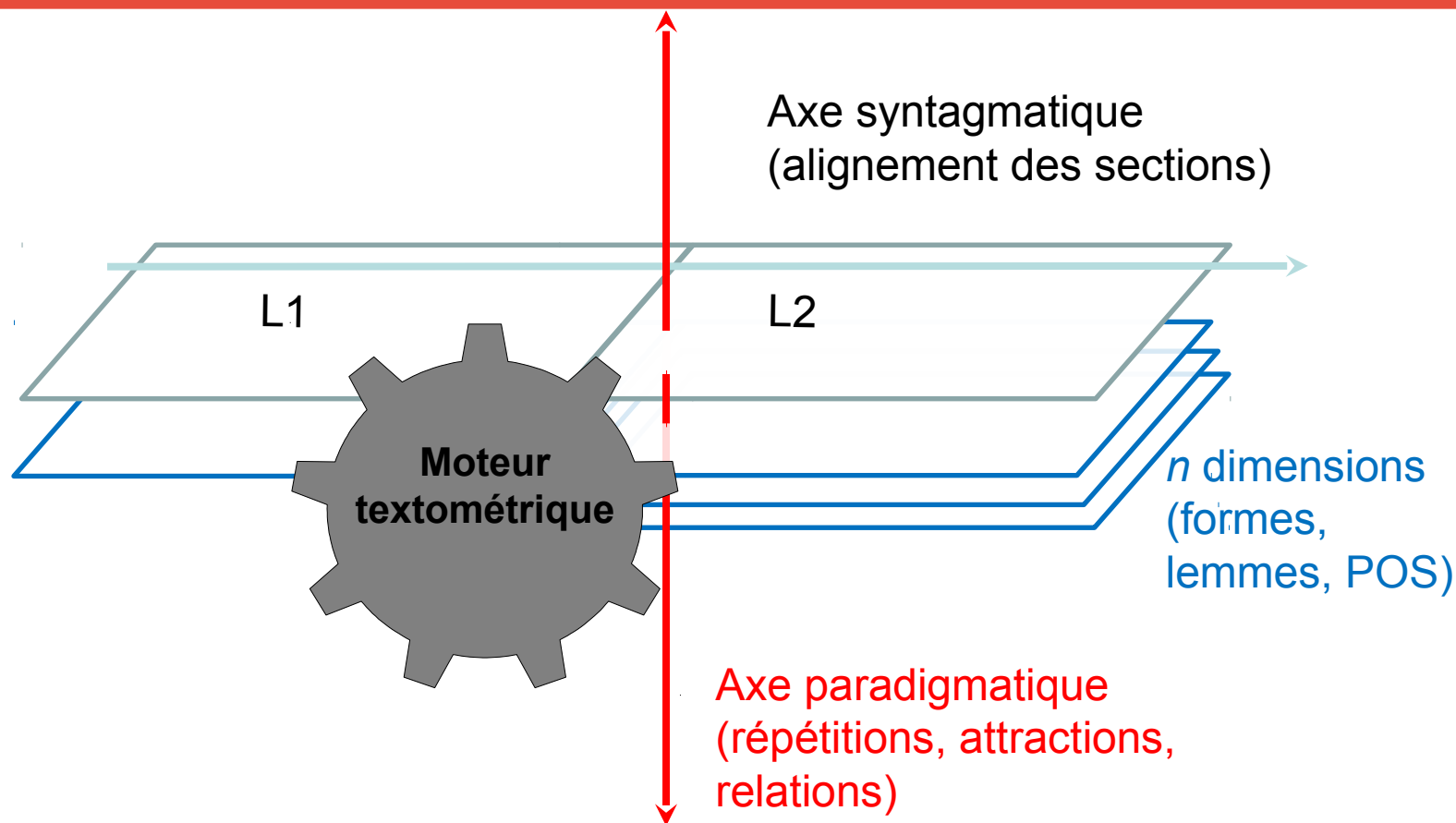
$K_{ij}$  : fréquence de l'unité  $j$  dans la partie  $i$

$F_i$  : fréquence de  $i$  dans le corpus

$t_j$  : taille de la partie  $j$



## 2. Bi-textes annotés : démo *iTrameur*



Zimina M., Fleury S. (2018). “Mémoire de traduction [MT] : approche paradigmatique. *Equivalences 45/1-2*, “Des unités de traduction à l’unité de traduction“.  
Sous la direction de C. Balliu, N. Froeliger, L. Hewson.

# 3. Données pluritextuelles multiannotées (démonstration *Trameur*)

Le texte dans le Trameur  
1. Une segmentation  
2. Systèmes de parties

on "découpe" des parts

Un texte est segmenté  
en unités  
(les mots en général)

Pour chaque mot

etc...

P.O.S

lemmes

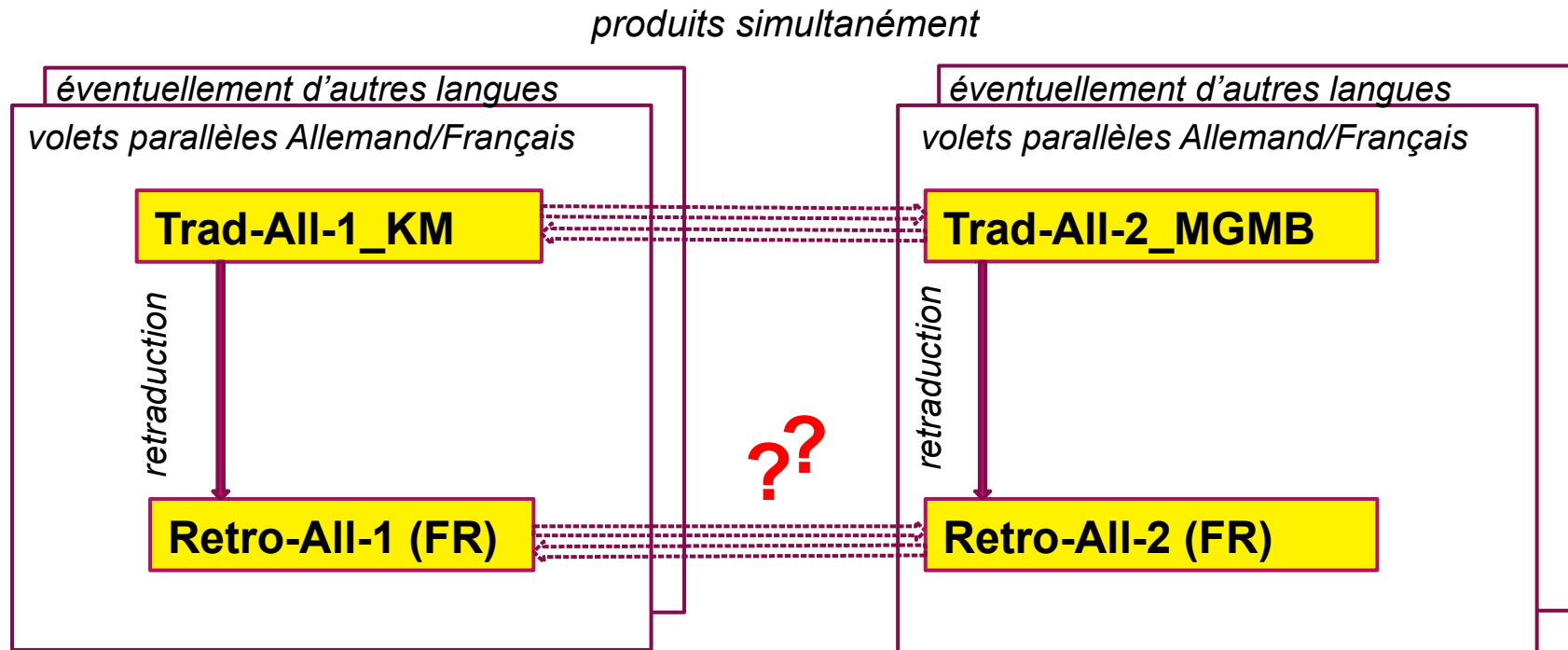
Formes

Zimina M., Fleury S.  
(2015)

« Perspectives de  
l'architecture Trame/  
Cadre pour les  
alignements  
multilingues. »  
NPSS, volume 11,  
numéro 1.

TEXTOMETRIE : compter des trucs dans des machins (trucs = mots..., machins = parties...)

# Extrait du corpus MACE : Prose PROCHE AFRIQUE (ALL\_FR)

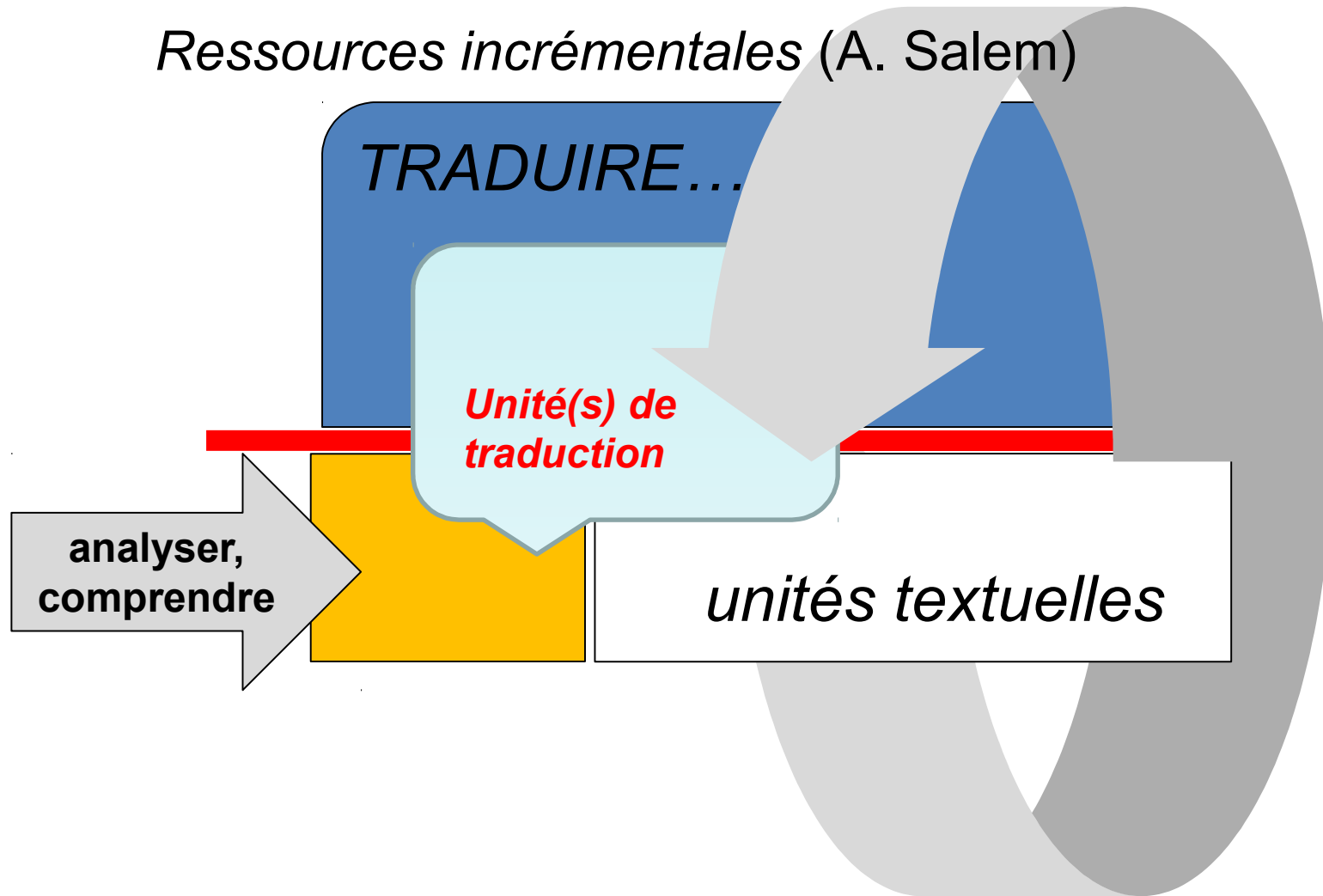


lectures partagées ? mondes discursifs communs ?

**=> proximités segmentales**

# Navigation textométrique

*Ressources incrémentales (A. Salem)*



# Références, bases textométriques multilingues...

<http://www.tal.univ-paris3.fr/trameur/>

<http://www.tal.univ-paris3.fr/trameur/iTrameur/>

[http://www.clillac-arp.univ-paris-diderot.fr/user/maria\\_zimina](http://www.clillac-arp.univ-paris-diderot.fr/user/maria_zimina)

**MERCI**